1-16-2013

Examining the Effects of Variation in Emotional Tone of Voice on Spoken Word Recognition

Maura L. Krestar

Conor T. McLennan
Cleveland State University, c.mclennan@csuohio.edu

Recommended Citation

Krestar, M. L., & McLennan, C. T. (2013). Examining the effects of variation in emotional tone of voice on spoken word recognition. The Quarterly Journal of Experimental Psychology, 66(9), 1793-1802.

Examining the effects of variation in emotional tone of voice on spoken word recognition

Maura L. Krestar and Conor T. McLennan

Emotional tone of voice (ETV) is essential for optimal verbal communication. Research has found that the impact of variation in nonlinguistic features of speech on spoken word recognition differs according to a time course. In the current study, we investigated whether intratalker variation in ETV follows the same time course in two long-term repetition priming experiments. We found that intratalker variability in ETVs affected reaction times to spoken words only when processing was relatively slow and difficult, not when processing was relatively fast and easy. These results provide evidence for the use of both abstract and episodic lexical representations for processing within-talker variability in ETV, depending on the time course of spoken word recognition.

Keywords: Spoken word recognition; Emotional tone of voice; Indexical specificity effects; Variability; Time course.

Nonlinguistic properties of speech, including emotional tone of voice (ETV), are essential for optimal verbal communication. Typically characterized by variations in prosody, loudness, pitch, syllable duration, and voice quality, a speaker's affective state directly influences ETV, providing listeners with information about the speaker's emotional and attitudinal state (Cutler, Dahan, & Donselaar, 1997). There is some evidence to suggest that emotional information is processed differently from lexical information, as specific brain areas tend to exhibit stronger responses to emotional speech than to neutral speech, which cannot be accounted for by single acoustic parameters (Ethofer et al., 2012). Although ETV has a powerful influence on communication, the way ETV is processed during spoken word recognition is not fully understood.

Researchers have adopted two general theories explaining the role of nonlinguistic variation during spoken word recognition. In extremely abstract theories, nonlinguistic properties are characterized as noise that must be stripped away, or normalized, to reveal the underlying phonological content of a spoken word, which then gets mapped onto form-based lexical representations. After normalization, the noise is discarded and does not form part of the lexical representation. Research supporting normalization has found that variation in nonlinguistic properties of speech has processing consequences for spoken word recognition (e.g., Mullennix, Pisoni, & Martin, 1989). Alternatively, episodic theories argue against normalization, asserting that nonlinguistic features of speech, including ETV, are encoded as part of form-based lexical representations and are thus preserved during spoken word recognition. Research in support of episodic representations has found that nonlinguistic information gets stored as part of the lexical representation and impacts later perceptual processing (e.g., Church & Schacter, 1994). There is an emerging consensus that a complete understanding of spoken word recognition requires recognition of both abstract and episodic

representations. As a result, less extreme views, referred to as "weakly abstract" or "weakly episodic" have emerged to encompass the ways we use both types of representations during spoken word recognition (Tenpenny, 1995).

Recent research has found that the time course of spoken word recognition impacts whether abstract or episodic representations dominate processing (M^{c}Lennan & Luce, 2005). Specifically, variation in talker voice impacted word recognition (specificity effects) when processing was relatively slow, but not when processing was relatively fast. Further research has supported the time-course hypothesis (however, see Orfanidou, Davis, Ford, & Marslen-Wilson, 2011). First, Mattys and Liss (2008) found that specificity effects emerged when participants responded slowly to degraded dysarthritic speech, but not when participants responded quickly to speech produced by healthy individuals. Second, Vitevitch and Donoso (2011) recently found that participants detected a change in talkers less often (i.e., a greater occurrence of change deafness) in an easy lexical decision task when responding quickly. Finally, M^{c}Lennan and González (2012) observed greater effects of variation in talker voice when participants responded slowly to words in a foreign accent than when they responded quickly to words in a native accent. These findings primarily support weakly episodic theories of spoken word recognition; however, they also make it clear that abstract theories should not be rejected entirely, as we use abstract representations early during spoken word recognition.

Previous research has also shown that variation in ETV affects the mechanisms through which humans process linguistic features of speech. Mullenix, Bihon, Bricklemyer, Gaston, and Keener (2002) investigated this relationship using a short-term speeded classification task. Participants responded more slowly when either the talkers or the ETV mismatched than when they matched. These results support episodic theories and suggest that participants should take longer to respond when the ETV mismatches from one time to another than when the ETV matches from one time to another in a long-term priming task.

The current study informs theories of spoken word recognition by addressing the long-lasting effects that variability has on the nature of representations underlying spoken language, as opposed to previous research regarding how such variability affects the processing of spoken words (e.g., Mullennix, Bihon, Bricklemyer, Gaston, & Keener, 2002). Using long-term priming in two experiments, the current study is the first to investigate whether intratalker variability in ETV affects listeners' perception of spoken words at different times during perceptual processing.

We used the long-term repetition priming paradigm and a lexical decision task in two experiments to investigate specificity effects of ETV. Experiments 1 and 2 differed only in their sets of nonwords. In Experiment 1, the nonwords were unwordlike (*yeesh-geesh*), making the discrimination between words and nonwords relatively easy and processing of all items, including the experimental words, relatively fast. In Experiment 2, the nonwords were wordlike (*bacov*), making the discrimination relatively difficult and processing relatively slow. The central research question under investigation was whether mismatches in ETV take time to influence spoken word recognition, as predicted by the time-course hypothesis. Therefore, we expected an attenuation (or absence) of specificity effects of ETV in Experiment 1, when processing was relatively fast, and greater specificity effects in Experiment 2, when processing was relatively slow. These results would indicate that abstract representations are more likely to dominate early during spoken word recognition, and episodic representations relatively late.

Alternatively, it is possible that no specificity effects will emerge. Such results would suggest that more abstract representations that are void of ETV information dominate throughout the duration of spoken word recognition, consistent with abstract theories. A final possibility is that specificity effects will emerge in both the easy and the hard experiments, implying that episodic representations that include ETV information dominate throughout the duration of spoken word recognition. Thus, although our prediction at the outset of the study is based on the time-course hypothesis, any of the three patterns would inform current theories of spoken word recognition.

EXPERIMENT 1: EASY DISCRIMINATION

Method

Participants

Seventy-five right-handed native English speakers with no reported history of speech or hearing disorders were recruited from Cleveland State University and received partial or extra credit for participating.

Materials

Auditory stimuli consisted of 12 bisyllabic experimental words spoken in a frightened emotional tone and 12 bisyllabic experimental words spoken in a sad emotional tone; 12 bisyllabic nonwords spoken in a frightened emotional tone and 12 bisyllabic nonwords spoken in a sad emotional tone; and 8 bisyllabic control items (4 words, 4 nonwords).

All words and nonwords were taken from Experiment 1 of M^{c}Lennan and Luce (2005). However, new auditory stimuli were recorded with a different speaker in two emotional tones of voice. To make the word–nonword discrimination relatively easy, the nonwords were unwordlike. The nonwords for Experiment 1 were created by using sequences with low phonotactic probability, determined by positional segment frequency and biphone frequency. All words and nonwords were spoken by the same speaker in both frightened and sad ETVs.

The mean log frequency of occurrence for the experimental stimuli was 0.79 (Kučera & Francis, 1967). Upon initial recording, the experimental words spoken in a frightened ETV had a shorter duration than the experimental words spoken in a sad ETV (frightened $M = 733$ ms, $SE = 33$; sad $M = 955$ ms, $SE = 36$), $t(22) = -4.551$, $p < .001$, Cohen's $d = -1.859$,[1] which is typical of emotional speech in English. To avoid confounding durational differences with reaction times (RTs), we equalized the durations of all word/nonword items by expanding the sad version of the item and compressing the frightened version of the item to match their mean duration. However, because a large amount of information contained in ETV is conveyed through the rate and duration of a spoken word, it is possible that artificially manipulating (equalizing) the durations of the stimuli could have unintended effects. Consequently, another condition was added in which the stimulus durations were "natural" (i.e., unedited stimuli). Thirty-seven of the participants heard equalized stimuli, and 38 heard natural stimuli. Prior to conducting the main experiments, two separate screenings confirmed the intelligibility of the natural and equalized word stimuli and the distinguishability of the ETVs. For the natural and equalized versions of all experimental words, at least 8 out of 10 screening participants correctly shadowed each word, and at least 8 out of 10 separate screening participants indicated the intended ETV for each stimulus word. A mean of 98% of participants correctly shadowed each natural word, and 96% identified the intended ETV of each natural word. A mean of 98% of participants correctly shadowed each equalized word, and 97% identified the intended ETV of each equalized word.

Recording auditory stimuli

We recorded auditory stimuli in a sound-attenuated room using Praat software (Boersma & Weenink, 2006). A female speaker of a Midwestern dialect was paid $25 to portray words and nonwords in both frightened and sad ETVs. To achieve the appropriate emotional tones, she imitated the tone of a fictional character in a brief, written emotional situation (Leinonen, Hiltunen, Linnankoski, & Laasko, 1997). Stimuli were edited into individual files and stored for later playback.

Design

Auditory stimuli spoken in frightened and sad ETVs were presented in two blocks: a prime followed by a target. Between the prime and target blocks, participants completed math problems for three to five minutes. For each block, half the stimuli were spoken in a frightened ETV

[1] Cohen's d statistics were calculated for within-participant data using an online effect size calculator (Cognitive Flexibility Laboratory, 2008). The typical effect size interpretations for Cohen's d are 0.2 = small; 0.5 = medium; 0.8 = large.

and half in a sad ETV. Primes were matched, mismatched, or unrelated to targets. ETVs of matched primes and targets were identical ($circus_{frightened}$, $circus_{frightened}$). ETVs of mismatched primes and targets differed ($circus_{sad}$, $circus_{frightened}$). Both prime and target blocks consisted of 24 stimuli: 12 words and 12 nonwords. The prime block included 8 experimental words, 8 nonwords, and 8 control stimuli (4 words, 4 nonwords). Control words and nonwords, by definition, were not repetition trials. The target block consisted of 12 experimental words and 12 nonwords. In the target block, 8 stimuli matched, 8 mismatched, and 8 were unprimed. Although preparation of the nonwords and their rotation through the various conditions paralleled the real word experimental stimuli, the nonwords and unrelated control stimuli were simply fillers; the focus of the manipulations and analyses was limited to the experimental words.

Orthogonal combination of prime (match, mismatch, unprimed) and target (frightened, sad) resulted in six completely within-participants conditions for each of the two between-participant stimulus conditions (natural, equated). Across participants, each frightened and sad item appeared in every possible condition. However, no single participant heard more than one version of a given word within a block.

Procedure

After providing informed consent, participants performed a lexical decision task in which they decided as quickly and accurately as possible whether the item they heard was a real word or a nonword by pressing a green button for word on the right or a red button for nonword on the left on a response box. After the participant responded, the next trial was initiated. After 5,000 ms, the computer automatically recorded an error and presented the next trial.

In both the prime and target blocks, stimuli were presented binaurally over headphones. Stimulus presentation within each block was randomized for each participant. RTs were measured from the onset of the stimulus to onset of the button press response.

Results

First, separately for the equated and natural durations stimulus conditions, we excluded participants whose overall mean percentage correct (PC) fell two standard deviations below the grand mean, resulting in the elimination of two participants in the equated condition and two in the natural condition. Next, we were prepared to replace missing RTs due to errors in both trials in a given condition with the mean of the corresponding condition; however, no such replacements were necessary in either stimulus condition. Finally, we replaced RTs more than two standard deviations beyond each condition mean with the mean of the corresponding condition, resulting in the replacement of six cells in the equated condition and eight cells in the natural condition (i.e., less than 4% of the cells).

Mean PCs and RTs as a function of condition, magnitude of specificity (MOS), and magnitude of priming (MOP) are shown in Tables 1 and 2, respectively. MOS is the difference in RT or PC between the match and mismatch conditions. MOP_{match} is the difference between the match and unprimed conditions. $MOP_{mismatch}$ is the difference between the mismatch and unprimed conditions.

Overall, RTs to primed (match and mismatch) conditions were shorter on average than RTs to the unprimed condition, primed $M = 945$ ms, unprimed $M = 1{,}020$ ms, $t(71) = -5.560$, $p < .001$, Cohen's $d = -0.0619$. In addition, PCs to primed conditions were higher than PCs to the unprimed condition; however, this difference was not significant, primed $M = 95\%$, unprimed $M = 93\%$, $t(71) = 1.403$, $p = .165$, Cohen's $d = 0.213$.

Prime (match, mismatch, unprimed) × ETV (sad, frightened) × Stimulus (natural, equated) participant analyses of variance (ANOVAs) were performed on mean RTs to correct responses and PCs for the experimental words in the target block. Accuracy was 94% overall. There was a significant main effect of stimulus; participants had higher PCs in the natural stimulus condition than in the equated, $M = 96$ and 92%, respectively, $F(1, 59) = 14.643$, $p < .001$, $\eta_p^2 = .199$. Accuracy

Table 1. *Mean percentage correct to experimental words as a function of prime, MOS, and MOP*

		Match		*Mismatch*		*Unprimed*				
Difficulty condition	n	*PC*	SE	*PC*	SE	*PC*	SE	*MOS*	MOP_{match}	$MOP_{mismatch}$
Easy	71	94	1	96	1	92	1	−2	2	4
Hard	72	96	2	92	2	90	2	4	6	2
Total	143	95	2	94	2	91	2	1	4	3

Note: PC = percentage correct; *SE* = standard error of the mean; MOS = magnitude of specificity (match minus mismatch); MOP_{match} = magnitude of priming for the match condition (match minus unprimed); $MOP_{mismatch}$ = magnitude of priming for the mismatch condition (mismatch minus unprimed).

did not yield any additional significant effects. Although responses to the nonwords were not the focus of the current study, the overall mean RT and PC for the nonword stimuli were 1,200 ms and 96%, respectively (*SE*s = 23 ms and 1%), indicating that participants were both fast and accurate in response to the nonwords.[2]

For RTs, there was no main effect of stimulus, $F(1, 59) = 0.184$, $p = .670$, $\eta_p^2 = .003$. Stimulus interacted with ETV, $F(1, 59) = 29.112$, $p < .001$, $\eta_p^2 = .330$. RTs to equalized stimuli, as expected, differed less between ETVs than did RTs to natural stimuli. For equalized stimuli, mean RTs to stimuli in frightened and sad ETVs were 944 and 1,001 ms, respectively. For natural stimuli, mean RTs to stimuli in frightened and sad ETVs were 876 and 1,047 ms, respectively. Crucially, stimulus did not interact with prime, indicating that the priming effects were equivalent in the two stimulus conditions (natural, equated), $F(2, 118) = 1.010$, $p = .367$, $\eta_p^2 = .017$. Furthermore, the same pattern of results was obtained in both stimulus conditions. Consequently, all subsequent analyses focus on data collapsed over the two stimulus conditions.

There was a main effect of ETV; participants responded to frightened items significantly more quickly than to sad items, frightened $M =$ 913 ms, sad $M = 1{,}026$ ms, $F(1, 59) = 113.839$, $p < .001$, $\eta_p^2 = .659$. PCs did not differ for items in frightened and sad ETVs, frightened $M =$ 95%, sad $M = 94\%$, $F(1, 59) = 1.186$, $p = .281$, $\eta_p^2 = .020$.

Of primary interest was the main effect of prime, which was significant, $F(2, 118) = 28.361$, $p < .001$, $\eta_p^2 = .325$. As expected, prime and ETV did not interact, $F(2, 118) = 1.622$, $p = .202$, $\eta_p^2 = .027$. Planned comparisons based on the main effect of prime revealed significant differences between the match and unprimed conditions (MOP_{match}) and between the mismatch and unprimed conditions ($MOP_{mismatch}$), indicating that both the match and mismatch conditions served as effective primes (both $ps < .001$). As expected, there was no difference between the match and mismatch conditions (i.e., no MOS; $p > .99$).

Discussion

Both matched and mismatched primes significantly facilitated lexical decision responses. Moreover, mismatched primes in the easy lexical decision experiment facilitated responses to targets as much as primes matched on ETV. These results

[2] Traditional item analyses with items as random factors are inappropriate for the current experiments, as we carefully selected stimuli on the basis of variables known to affect the dependent variables under investigation. Furthermore, the design used counterbalanced lists such that each item appeared in every condition (Raaijmakers, Schrijnemakers, & Gremmen, 1999). Furthermore, long-term repetition priming paradigms limit the number of items in a within-participants manipulation because increasing items tends to decrease the likelihood of obtaining long-term repetition priming effects, in turn decreasing power as a consequence of having few items (M^{c}Lennan & Luce, 2005). Therefore, two dummy variables representing allocation of participants to experimental lists were included in the ANOVAs. Because these dummy variables were included solely to reduce the estimate of random variation (see Gaskell & Dumay, 2003; Pollatsek & Well, 1995), their effects are not reported.

Table 2. *Mean reaction times to experimental words as a function of prime, MOS, and MOP*

		Match		*Mismatch*		*Unprimed*				
Difficulty condition	n	*RT*	SE	*RT*	SE	*RT*	SE	*MOS*	MOP_{match}	$MOP_{mismatch}$
Easy	71	945	15	939	13	1,020	15	6	−75	−81
Hard	72	1,031	17	1,071	14	1,127	20	−40	−96	−56
Total	143	988	16	1,005	14	1,074	18	−17	−86	−69

Note: RT = reaction time, in ms; *SE* = standard error of the mean; MOS = magnitude of specificity (match minus mismatch); MOP_{match} = magnitude of priming for the match condition (match minus unprimed); $MOP_{mismatch}$ = magnitude of priming for the mismatch condition (mismatch minus unprimed).

are consistent with the previously discussed time-course hypothesis: When processing was fast (as a result of the easy discrimination allowed by the unwordlike nonwords), indexical specificity effects of ETV did not emerge.

Experiment 2 was conducted to test the hypothesis that when processing is slowed by the use of wordlike nonwords, indexical specificity effects of ETV should emerge with the same experimental words as those used in Experiment 1.

EXPERIMENT 2: HARD DISCRIMINATION

This experiment is essentially a replication of Experiment 1, except for the nonwords. Instead of using unwordlike nonwords, we used wordlike nonwords in Experiment 2. This change was expected to slow participants' processing of all items, including the experimental words. Therefore, we predicted that we would obtain indexical specificity effects for variability in ETV.

Method

Participants

Seventy-five different participants were recruited from the same population and met the same criteria as those in Experiment 1.

Materials

The stimuli consisted of (a) the same 12 bisyllabic spoken experimental words as those used in Experiment 1, (b) 12 new spoken bisyllabic nonwords, and (c) eight bisyllabic spoken control items (4 words, 4 nonwords). To increase the difficulty of the word–nonword discrimination task, the nonwords were wordlike. Wordlike nonwords were taken from M^{c}Lennan and Luce (2005), which were created by changing the endings of real words so that they became nonwords (e.g., *bygone, bygups*). These nonwords were spoken by the same speaker in both frightened and sad ETVs. The stimuli were recorded in a sound-attenuated room by the same speaker as Experiment 1. All stimuli were edited into individual files and stored on computer disk for later playback. Thirty-eight of the participants heard equalized stimuli, and 37 heard natural stimuli.

Design and procedure

The design and procedure were identical to those described in Experiment 1.

Results

Again, separately for the equated and natural conditions, we excluded participants whose overall mean percentage correct (PC) fell two standard deviations below the grand mean, resulting in the elimination of one participant in the equated condition and two in the natural condition. Next, we replaced missing RTs due to errors in both trials in a given condition with the mean of the corresponding condition, resulting in the replacement of four cells in the equated condition and two cells in the natural condition (i.e., less than 1.5% of the cells). Finally,

we replaced RTs more than two standard deviations beyond each condition mean with the mean of the corresponding condition, resulting in the replacement of 10 cells in the equated condition and 9 cells in the natural condition (i.e., less than 4.5% of the remaining cells).

Overall, RTs to primed conditions were shorter on average than RTs to the unprimed condition, primed $M = 1{,}053$ ms, unprimed $M = 1{,}139$ ms, $t(70) = -6.242$, $p < .001$, Cohen's $d = -0.686$. In addition, PCs to primed conditions were higher than PCs to the unprimed condition; however, this difference was not significant, primed $M = 94\%$, unprimed $M = 90\%$, $t(70) = 1.703$, $p = .093$, Cohen's $d = 0.266$.

Prime (match, mismatch, unprimed) × ETV (sad, frightened) × Stimulus (natural, equated) participant ANOVAs were performed on mean RTs to correct responses and PCs for the experimental words in the target block. Accuracy was 92% overall. There was a significant main effect of stimulus; participants had higher PCs in the natural stimulus condition than in the equated, $M = 98$ and 89%, respectively, $F(1, 63) = 27.875$, $p < .001$, $\eta_p^2 = .307$. A marginally significant main effect on PCs emerged for ETV, $F(1, 63) = 3.984$, $p = .050$, $\eta_p^2 = .059$. Mean PC was 94% for words in a frightened ETV and 90% for words in a sad ETV. For PCs, there was no main effect of prime, $F(2, 126) = 1.492$, $p = .229$, $\eta_p^2 = .023$. However, a marginally significant Prime × Stimulus interaction emerged, $F(2, 126) = 2.733$, $p = .069$, $\eta_p^2 = .042$, such that the greater mean PC for the match condition than for the mismatch condition only emerged in the equated stimulus condition.

Mean PCs as a function of prime type are reported in Table 1. The overall mean RT and PC for the nonword stimuli were 1,242 ms and 89%, respectively (SEs = 27 ms and 1%).

For RTs, there was no main effect of stimulus, $F(1, 63) = 0.782$, $p = .380$, $\eta_p^2 = .012$. Stimulus interacted with ETV, $F(1, 63) = 30.080$, $p < .001$, $\eta_p^2 = .323$. Again, RTs to equalized stimuli, as expected, differed less between ETVs than did RTs to natural stimuli. For equalized stimuli, mean RTs to stimuli in frightened and sad ETVs were 1,050 and 1,078 ms, respectively. For natural stimuli, mean RTs to stimuli in frightened and sad ETVs were 996 and 1,185 ms, respectively. Crucially, stimulus did not interact with prime, indicating that the priming effects were equivalent in the two stimulus conditions (natural, equated), $F(2, 126) = 0.653$, $p = .522$, $\eta_p^2 = .010$. Furthermore, the same pattern of results was obtained in both stimulus conditions. Consequently, all subsequent analyses focus on data collapsed over the two stimulus conditions.

There was a main effect of ETV; frightened items were again responded to more quickly than sad items, frightened $M = 1{,}033$ ms, sad $M = 1{,}130$ ms, $F(1, 63) = 59.634$, $p < .001$, $\eta_p^2 = .486$. Again, a main effect of prime emerged, $F(2, 126) = 13.323$, $p < .001$, $\eta_p^2 = .175$. Again, prime and ETV did not interact, $F(1, 126) = 1.676$, $p = .191$, $\eta_p^2 = .026$. Planned comparisons based on the significant main effect of prime revealed the predicted differences between the match and unprimed conditions, $p < .001$, and between the mismatch and unprimed conditions, $p = .002$, indicating that, as in Experiment 1, both the match and mismatch conditions served as effective primes. More importantly, there was also a significant difference between the match and mismatch conditions, $p = .036$, indicating that the match condition served as a more effective prime than the mismatch condition.

Discussion

Both matched and mismatched primes produced facilitative effects on lexical decision responses. However, the difference between the matched and mismatched conditions demonstrates that words matching in ETV served as more effective primes than did primes mismatched on ETV. The pattern is consistent with our time-course predictions: When processing was relatively slow in Experiment 2, specificity effects of ETV emerged. In contrast, when processing was fast in Experiment 1, indexical specificity effects of ETV did not emerge. Consequently, these results provide further support for the general hypothesis that time course is an important factor in

determining the role that indexical variability plays in spoken word recognition.

COMBINED ANALYSIS OF EXPERIMENTS 1 AND 2

RTs to correct responses to experimental target words in the hard condition were significantly longer than those in the easy condition, $t(140) = -6.588$, $p < .001$, Cohen's $d = -1.0944$ (easy $M =$ 970 ms, hard $M = 1{,}082$ ms), indicating that the difficulty manipulation was successful. The same pattern emerged in PCs; participants made more errors in the hard condition than in the easy condition; however, this difference was not significant, $t(140) = 1.155$, $p = .250$, Cohen's $d = 0.193$ (easy $M = 94\%$, hard $M = 92\%$). For nonwords (mean PCs and RTs are reported in the Results sections), participants did not significantly differ in their RTs across experiments, $t(140) = -1.152$, $p = .251$, Cohen's $d = -0.194$. However, participants were significantly less accurate in the hard lexical decision task (Experiment 2) than in the easy task (Experiment 1), $t(140) = 3.960$, $p < .001$, Cohen's $d = 0.669$.

Finally, we found further support for the time-course hypothesis by testing the difference in MOS in RT between Experiments 1 and 2. MOS in the hard/slow task (Experiment 2) was significantly larger than MOS in the easy/fast task (Experiment 1), $t(140) = 2.708$, $p = .008$, Cohen's $d = 0.718$.

GENERAL DISCUSSION

The central research question was whether mismatches in ETV take time to influence spoken word recognition, as predicted by the time-course hypothesis. We predicted that specificity effects of ETV would emerge when processing was slow, but not when processing was fast.

Regardless of whether the stimulus durations were natural or equalized across ETVs, both matched and mismatched primes facilitated lexical decision responses relative to words that were not repeated from the prime block to the target block (i.e., words in the unprimed condition). The results indicated that patterns of facilitative effects differed depending on the speed of responses. In Experiment 1, when processing was fast/easy, words matched and mismatched in ETV were equally effective primes. In Experiment 2, when processing was slow/hard, primes matched in ETV were more effective than were mismatched primes.

The current results have important implications for spoken word recognition models that account for effects of nonlinguistic variation in speech. In particular, the results indicate that abstract representations tend to play a role early during spoken word recognition, and episodic representations affect perceptual processing relatively late. Thus, listeners use both abstract and episodic representations, and which type of representation listeners use depends (at least in part) on the time course of processing. This time-course pattern occurs even when the speech signal consists of within-talker variability in ETV.

We are not arguing that time course is the only aspect of spoken word processing that can account for the effects of variation in ETV and other nonlinguistic features of speech. Future research should consider other factors that may also influence the nature of the representation of words during spoken word recognition. For example, attention to a particular nonlinguistic feature like ETV may allow episodic representations to dominate, even when processing is relatively fast. Initial support for attentional modulation of the time course of specificity effects has been reported (Theodore & Blumstein, 2011). It is also possible that emotional words draw listeners' attention to the ETV, or other nonlinguistic properties of the input. This possibility is supported by research demonstrating that ETV facilitates linguistic processing when words with emotional meanings are spoken in congruent emotional tones (Nygaard & Queen, 2008; Wurm, Vakoch, Strasser, Calin-Jageman, & Ross, 2001). Similarly, a listener's mood could bias attention toward a congruent/incongruent ETV, influencing which type of representation is likely to dominate recognition. In addition, age differences in attentional biases to positive and

negative stimuli (Thomas, 2006) might influence which type of representation is accessed when recognizing words spoken in positive or negative emotional tones.

Although we have interpreted our results in the context of the time-course hypothesis, it is possible that the use of different types of nonwords affects more than just task difficulty and thus the time course of processing. For example, it is possible that participant strategies might shift as a function of the nonword manipulation. Indeed, one possibility, consistent with the attention-based explanation mentioned above, is that participants devote more attentional resources to processing the stimuli when the task is hard, and thus the attentional differences may underlie the effect, and not time, per se.

Although it was not a central focus of the current study, it is interesting that participants were more accurate when responding to words spoken in a frightened ETV than to words spoken in a sad ETV, even when durations of the stimuli were equalized across ETVs. Recent research has provided some evidence for the evolutionary relevance of fear over other emotions (Brosch, Sander, Pourtois, & Scherer, 2008). It is possible that reacting more accurately to someone speaking in a frightened ETV aided survival relative to reacting to someone speaking in a sad ETV.

In conclusion, the present findings provide important new information that adds to our understanding of lexical representations involved in spoken word recognition. It is now clear that theories of spoken word recognition need to account for the effects of intratalker variation in ETV on the speech signal over time.

REFERENCES

Boersma, P., & Weenink, D. (2006). Praat: doing phonetics by computer (Version 4.5.08) [Computer program].

Brosch, T., Sander, D., Pourtois, G., & Scherer, K. R. (2008). Beyond fear: Rapid spatial orienting toward positive emotional stimuli. *Psychological Science, 19,* 362–370.

Church, B. A., & Schachter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(3), 521–533.

Cognitive Flexibility Laboratory (2008). *Effect size calculator,* Retrieved March 5, 2012, from http://www.cognitiveflexibility.org/effectsize

Cutler, A., Dahan, D., & Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40,* 141–201.

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., & Vuilleumier, P. (2012). Emotional voice areas: Anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. *Cerebral Cortex, 22,* 191–200.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89,* 105–132.

Kučera, H., & Francis, W. (1967). *Computational analysis of present day American English.* Providence, RI: Brown University Press.

Leinonen, L., Hiltunen, T., Linnankoski, I., & Laasko, M. (1997). Expression of emotional-motivational connotations with a one-word utterance. *Journal of the Acoustical Society of America, 102,* 1853–1863.

Mattys, S. L., & Liss, J. M. (2008). Building models of spoken-word recognition: When there is as much to learn from natural "oddities" as artificial normality. *Perception and Psychophysics, 70*(7), 1235–1242.

McLennan, C. T., & González, J. (2012). Examining talker effects in the perception of native- and foreign-accented speech. *Attention, Perception, and Psychophysics, 74,* 824–830.

McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 306–321.

Mullennix, J. W., Bihon, T., Bricklemyer, J., Gaston, J., & Keener, J. M. (2002). Effects of variation in emotional tone of voice on speech perception. *Language and Speech, 45,* 255–283.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85,* 365–378.

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance, 34*(4), 1017–1030.

Orfanidou, E., Davis, M. H., Ford, M. A., & Marslen-Wilson, W. D. (2011). Perceptual and response components in repetition priming of spoken words and pseudowords. *The Quarterly Journal of Experimental Psychology, 64*(1), 96–121.

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 785–794.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41,* 416–426.

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review, 2,* 339–363.

Theodore, R. M., & Blumstein, S. E. (2011). Attention modulates the time-course of talker-specificity effects in lexical retrieval. *Journal of the Acoustical Society of America, 130,* 2442–2442.

Thomas, R. C. (2006). The influence of emotional valence on age differences in early processing and memory. *Psychology and Aging, 21*(4), 821–825.

Vitevitch, M. S., & Donoso, A. (2011). Processing of indexical information requires time: Evidence from change deafness. *The Quarterly Journal of Experimental Psychology, 64*(8), 1484–1493.

Wurm, L. H., Vakoch, D. A., Strasser, M. R., Calin-Jageman, R., & Ross, S. E. (2001). Speech perception and vocal expression of emotions. *Cognition and Emotion, 15*(6), 831–852.